TENTAMENSKOD:_____

ÖREBRO UNIVERSITET
INSTITUTIONEN FÖR
NATURVETENSKAP OCH TEKNIK

Kursens namn/kurskod	**Numeriska Metoder för civilingenjörer DT508G**
Examinationsmomentets namn/provkod	**Teori, 3,5 högskolepoäng (Provkod: A001)**
Datum	2020-01-03
Tid	Kl. 14:15 – 19:15

Tillåtna hjälpmedel	Skrivmateriel, formel blad och miniräknare med raderat minne.
Instruktion	Läs igenom alla frågor noga. Börja varje fråga på ett nytt svarsblad. Skriv bara på ena sidan av svarsbladet. Skriv tentamenskoden på varje svarsblad. Skriv läsligt! Det räcker att ange fem decimaler.
Viktigt att tänka på	Motivera väl, redovisa alla väsentliga steg, rita tydliga figurer och svara med rätt enhet. Lämna in i uppgiftsordning.
Ansvarig/-a lärare (ev. telefonnummer)	Danny Thonig (Mobil: 0727010037)
Totalt antal poäng	40
Betyg (ev. ECTS)	Skrivningens maxpoäng är 40. För betyg 3/4/5 räcker det med samt 20/28/35 poäng totalt. Detaljerna framgår av separat dokument publicerat på Blackboard.
Tentamensresultat	Resultatet meddelas på Studentforum inom 15 arbetsdagar efter tentadagen.
Övrigt	Lärare är inte på plats. Varsågod och ringa om du har frågar.

Lycka till!

1. Decide whether the following statements are true or false. Explain the reason for your answer:

a) Integrals on non-rectangular regions can not be numerically solved by the trapezoid rule. [1p]

b) Let f be a continuously differentiable function on $[a,b]$ and $x_0, x_1, \ldots, x_n \in [a,b]$ are distinct, i.e.,$x_i \neq x_j, i \neq j$. Then the Hermite polynomial that interpolates $(x_0, f(x_0)), \ldots, (x_n, f(x_n))$ has the degree n. [1p]

c) The false position method is a bisection method, where the midpoint is replaced by the secant. [1p]

d) The Hilbert matrix has a small condition number for large matrices. [1p]

e) Clamped cubic splines S are created by putting the first derivative of S_1 at x_1 and S_{n-1} at x_n to be arbitrary. [1p]

f) The machine error for double precession numbers is smaller than 10^{-20}. [1p]

g) Lets consider the problem $Ax = b$, where A is a $n \times n$ matrix. If A is strictly diagonal dominant, then for any vector b and starting guess x_0 the Gauss-Seidel method converges. [1p]

h) Adaptive integration methods improve the accuracy of the numerical integration method by adapting the step size according to the integrants slope. [1p]

i) The modified Newton method take into account the m-th derivative of the function f, say $f^{(m)}$, where $m > 2$. [1p]

j) The relative error between a real number x and its floating point representation is smaller than half of the machine precision. [1p]

Suggested Solution:

You get already [0.5p] if true or false Is correct.

a) [1p] False. Non-rectangular regions can be integrated as follows $\int_c^d \mathrm{d}x \int_{a(x)}^{b(x)} \mathrm{d}y f(x,y)$ and trapezoid method can be applied to all integrals.

b) [1p] False. The Hermit polynomial is of the order $2n+1$.

c) [1p] True. The false position method is a bisection method where the midpoint is calculated by $c = \dfrac{f(a)b - f(b)a}{f(a) - f(b)}$.

d) [1p] False. The Hilbert matrix is a bad conditioned matrix for large matrices.

e) [1p] True. $S_1'(x_1)$ and $S_{n-1}'(x_n)$ are set to user-specific values.

f) [1p] False. The machine precision for double precision numbers is $2^{-52} = 2.2 \cdot 10^{-16}$.

g) [1p] True. Theorem 2.11 of the book.

h) [1p] True. The bigger the slop and the relative change of $f(x)$ with x, the smaller has the be the step width to guaranty a preselected error.

i) [1p] False. In the modified Newton method, m is the multiplicity of the root of f and the iteration is $x_{i+1} = x_i - \dfrac{mf(x_i)}{f'(x_i)}$.

j) [1p] True. $\dfrac{|fl(x) - x|}{|x|} \leq \dfrac{\varepsilon_{mach}}{2}$

2. The two numbers x=10.4 and *y*=0.5 should be represented into a **single precision** language understandable for the computer.
   a) Find the binary $(x)_2$, floating-point $fl(x)$, and machine representation in single precision of the two given numbers x and y. [4p]
   b) Calculate the error (rounding and chopping error) of the floating-point representation of the two numbers x and y. [2p]
   c) Use the binary number representation of x and y, perform the operations listed in the following, and find again the decimal number of the result. [4p]
      (I) Addition, say $(x)_2 + (y)_2$
      (II) Bit-wise logical AND, say $(x)_2 \,.\, AND \,.\, (y)_2$
      (III) Bit-wise logical OR, say $(x)_2 \,.\, OR \,.\, (y)_2$
      (IV) Bit-wise logical XOR, say $(x)_2 \,.\, XOR \,.\, (y)_2$

   The logic table for AND, OR, and XOR can be found in the formula sheet.

Suggested Solution:

a)

[1p] for the solution path

[1p] for binary solution $(10.4)_2 = \ldots 1010.\overline{0110}$ and $(0.5)_2 = \ldots 0.1\overline{0000}$.

[1p] for floating point solution $fl(10.4) = 1.010\overline{0110} \times 2^3$ and $fl(0.5) = 1.\overline{0000} \times 2^{-1}$.

[1p] for machine representation

Let the machine representation be like $se_1e_2 \ldots e_8b_1b_2 \ldots b_{23}$ in single precision.

Shift of the exponent $(2^{n^{exp}-1} - 1) + p$, where $n^{exp} = 8$ and, consequently, $(2^{n^{exp}-1} - 1) = 127$.

For 10.4, $p = 3$ and thus $e_1e_2 \ldots e_8 = 10000010$.

$mach(10.4) = 0\,|\,10000010\,|\,01001100110011001100110$

For 0.5, $p = -1$ and thus $e_1e_2 \ldots e_8 = 01111110$

$mach(0.5) = 0\,|\,01111110\,|\,00000000000000000000000$

b)

[1p] for calculating the particular errors.

Express the floating point number in single precession.

$fl(10.4) = 1.01001100110011001100110\,|\,\overline{0110} \times 2^3 \approx 1.01001100110011001100110 \times 2^3$

Let error$_c$ and error$_r$ be the copping and the rounding error, respectively. Thus,

error$_c = .\,\overline{0110} \times 2^3 \times 2^{-23}$ and error$_r = 0$.

$fl(0.5) = 1.00000000000000000000000\,|\,\overline{0000} \times 2^{-1} \approx 1.00000000000000000000000 \times 2^{-1}$

Let error$_c$ and error$_r$ be the copping and the rounding error, respectively. Thus,

error$_c = 0$ and error$_r = 0$. 0.5 is represented exact.

c)

[1p] for calculating the errors

$1010.0\overline{1100}$

$0000.1\overline{0000}$.

I) ADD : $1010.1\overline{1100} = (10.9)_{10}$
II) AND : $0000.0\overline{0000} = (0)_{10}$
III) OR : $1010.1\overline{1100} = (10.9)_{10}$
IV) XOR : $1010.1\overline{1100} = (10.9)_{10}$

3. The population dynamics of three competing species $x, y,$ and z can be described by

$$x'(t) = x(t)\big[1 - x(t) - \alpha y(t) - \beta z(t)\big]$$

$$y'(t) = y(t)\big[1 - y(t) - \beta x(t) - \alpha z(t)\big]$$

$$z'(t) = z(t)\big[1 - z(t) - \alpha x(t) - \beta y(t)\big],$$

where α and β are parameters measuring the influence that the species have on each other. The stable solution ($x'(t) = y'(t) = z'(t) = 0$) of the scaled populations $\boldsymbol{x} = (x(t), y(t), z(t))$ can be found e.g. by using Newton and quasi-Newton method.

a) Describe the multivariable Newton method. What is the order and rate of convergence? When are quasi-newton methods more applicable than Newton's method? Describe Broyden's method 1, which is a quasi-newton method. What is the order of convergence for the quasi-Newton methods? [3p]

b) Calculate the Jacobian and the inverse of the Jacobian for the problem above. [3p]

c) If $\alpha = 0.5$ and $\beta = 0.25$, estimate a solution $\boldsymbol{x} = (x(t), y(t), z(t))$ in the set described by $0.0 \le x(t) \le 1$, $0.25 \le y(t) \le 1$, and $0.25 \le z(t) \le 1$. Here, calculate the first two step $\boldsymbol{x}_1, \boldsymbol{x}_2$ of Broyden's method 2? The initial value is $\boldsymbol{x}_0 = (0.0, 0.25, 0.25)$ and $A_0^{-1} = DF(\boldsymbol{x}_0)^{-1}$. To get the inversion, it is easier to calculate first $DF(\boldsymbol{x}_0)$ and invert this numerical matrix, e.g. by Gauss elimination. Compare your result with $DF(\boldsymbol{x}_0)$ and $DF(\boldsymbol{x}_0)^{-1}$ from the formula sheet. [4p]

Suggested Solutions:

a)

See text books, Sauer, Page 131-133 as well as Burden, Page 55-65 and Page 651 - 659.
[1p] Definition of newton method. $x_{i+1} = x_i - (DF(x_i))^{-1}F(x_i)$.
[1p] The order of convergence is two and the rate is similar to the on-dimensional case: $\frac{1}{2}(DF(x))^{-1}DDF(x)$, where $DF(x)$ is the Jacobian matrix and $DDF(x)$ is the derivative of the Jacobian.
[1p] Newton fails when Jacobian is nearly singular or unknown. Broyden method 1 approximates the Jacobian od the function F and iterates this matrix. Inversion is required.
[1p] The order of convergence is superlinear.

b)

We are looking for the problem, when
$0 = x\left[1 - x - \alpha y - \beta z\right]$
$0 = y\left[1 - y - \beta x - \alpha z\right]$
$0 = z\left[1 - z - \alpha x - \beta y\right],$
Where the right-hand side is $F(x) : R^3 \to R^3$ and $x = (x, y, z)$.

Thus, the Jacobian [1p] is

$$DF(x) = \begin{pmatrix} 1-2x-\alpha y-\beta z & -\alpha x & -\beta x \\ -\beta y & 1-2y-\beta x-\alpha z & -\alpha y \\ -\alpha z & -\beta z & 1-2z-\alpha x-\beta y \end{pmatrix}$$

To calculate the inverse [2p], Gauss elimination method can be applied. Or you use the relation from linear algebra $A^{-1} = \frac{1}{det(A)} adj(A)$.

c)

$\alpha = 0.5$ and $\beta = 0.25$, $x_0 = (0,0.25,0.25)$ and, thus,

$DF(x_0) = \begin{pmatrix} \frac{13}{16} & 0 & 0 \\ -\frac{1}{16} & \frac{3}{8} & -\frac{1}{8} \\ -\frac{1}{8} & -\frac{1}{16} & \frac{7}{16} \end{pmatrix}$ and using LU factorization and Gauss Elimination,

$DF(\boldsymbol{x}_0)^{-1} = \begin{pmatrix} \frac{16}{13} & 0 & 0 \\ \frac{22}{65} & \frac{14}{5} & \frac{4}{5} \\ \frac{2}{5} & \frac{2}{5} & \frac{12}{5} \end{pmatrix} = A_0^{-1}$ [1p] - but was also given in the appendix.

$F(x_0) = \begin{pmatrix} 0 \\ \frac{5}{32} \\ \frac{11}{64} \end{pmatrix}$, $A_0^{-1}F(x_0) = \left(0, \frac{23}{40}, \frac{19}{40}\right)$

The first iteration step is

$x_1 = x_0 - A_0^{-1}F(x_0)$ [1p]

For the second iteration we calculate first

$\boldsymbol{y}_1 = F(\boldsymbol{x}_1) - F(\boldsymbol{x}_0)$

$\boldsymbol{s}_1 = \boldsymbol{x}_1 - \boldsymbol{x}_0$

$A_1^{-1} = A_0^{-1} + \frac{(\boldsymbol{s}_1 - A_0^{-1}\boldsymbol{y}_1)\boldsymbol{s}_1^t A_0^{-1}}{\boldsymbol{s}_1^t A_0^{-1}\boldsymbol{y}_1}$ [1p]

And after

$x_2 = x_1 - A_1^{-1}F(x_1)$ [1p]

4. The fundamental theorem of interpolation says that each function can be approximated by a polynomial, such as the Lagrange polynomial. Based on this, solve the following part problems

   a) Derive Simpson's three-eights rule for the integral $\int_{x_0}^{x_3} f(x)\mathrm{d}x$ using equally spaced point x_0, $x_1 = x_0 + h$, $x_2 = x_0 + 2h$, and $x_3 = x_0 + 3h$. You can check the result with the one in the formula sheet. [5p]

   b) Calculate the error of Simpson's three-eights rule found above. What is the order of approximation. [2p]

   c) Based on the result of a) determine numerically the integral $\int_{-1/2}^{1/2} \tan(x)\mathrm{d}x$ and compare with the analytical solution, which is $\int \tan(x)\mathrm{d}x = -\ln(\cos(x))$. Note that you have to fix h first. [3p]

Suggested Solution:

a)

[1p] for (or Taylor expansion method. [1p] also when the way to the solution was mentioned)

$$f(x) = \sum_{k=0}^{3} L_k(x) f(x_k) + \frac{1}{4!} f^{(4)}(c(x)) \Pi_{k=0}^{3}(x - x_k).$$

Lets focus on $L_k(x)$

$$L_0(x) = \frac{(x - x_1)(x - x_2)(x - x_3)}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)}$$

$$L_1(x) = \frac{(x - x_0)(x - x_2)(x - x_3)}{(x_1 - x_0)(x_1 - x_2)(x_1 - x_3)}$$

$$L_2(x) = \frac{(x - x_0)(x - x_1)(x - x_3)}{(x_2 - x_0)(x_2 - x_1)(x_2 - x_3)}$$

$$L_3(x) = \frac{(x - x_0)(x - x_1)(x - x_2)}{(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)}$$

Thus

$$\int f(x)\mathrm{d}x = \sum_{k=0}^{3} (\underbrace{\int L_k(x)\mathrm{d}x) f(x_k)}_{*} + \frac{1}{24} \left(\int f^{(4)}(c(x)\mathrm{d}x) \right) \Pi_{k=0}^{3}(x - x_k) + \frac{1}{24} \left(\int \Pi_{k=0}^{3}(x - x_k)\mathrm{d}x \right) f^{(4)}(c(x))$$

Where the integral runs from x_0 to x_3

Lets consider first the constant terms in *

$k = 0$ it is $-\frac{f(x_0)}{6h^3}$,

$k = 1$ it is $\frac{f(x_1)}{2h^3}$,

$k = 2$ it is $-\frac{f(x_2)}{2h^3}$,
$k = 3$ it is $\frac{f(x_3)}{6h^3}$,

And the integrals for x_k, say $\int_{x_0}^{x_3} \mathrm{d}x \Pi_{i\neq k}^3 (x - x_i)$, are

$k = 0$ it is $4x\left(x^3 - \frac{4}{3}x^2\left(x_1 + x_2 + x_3\right) + 2x(x_1x_2 + x_1x_3 + x_2x_3) - x_1x_2x_3\right)\Big|_{x_0}^{x_3} =$

$-\frac{9h^4}{4}$,
$k = 1$ it is $\frac{9h^4}{4}$,
$k = 2$ it is $-\frac{9h^4}{4}$,
$k = 3$ it is $\frac{9h^4}{4}$.

Here we uses $x_1 = x_0 + h, x_2 = x_0 + 2h, x_3 = x_0 + 3h$ and, thus,

$$\frac{3h}{8}\left(f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)\right).$$

b)

The error is obtained by $\frac{1}{24}\left(\int_{x_0}^{x_3} \Pi_{k=0}^3 (x - x_k)\mathrm{d}x\right) f^{(4)}(c(x))$, which is

$-\frac{3h^5}{80} f^{(4)}(c(x))$.

c)

First, notice that $h = \frac{1}{3}$. For each result you get [1p].

$\tan(x_0) = -0.546302489843791$
$\tan(x_0 + h) = -0.168227218302242$
$\tan(x_0 + 2h) = 0.168227218302242$
$\tan(x_0 + 3h) = 0.546302489843791$

And, thus, $\int_{-1/2}^{1/2} \tan(x)\mathrm{d}x = 0$. This is also the result of the analytical solution $\ln(\cos(0.5) - \ln(\cos(-0.5) = 0$.